VID
CHALMERS TEKNISKA HÖGSKOLA
Nr 128

CHALMERS TEKNISKA HÖGSKOLA

CHALMERS UNIVERSITY OF TECHNOLOGY
GÖTEBORG
SWEDEN

STATISTICAL TESTS OF COMPOSITE HYPOTHESES AND THEIR ASYMPTOTIC PROPERTIES

STURE HOLM

DEPARTMENT OF MATHEMATICS

1973

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41544711_0016

Statistical tests of composite hypotheses
and their asymptotic properties
by
Sture Holm

Akademisk avhandling som med tillstånd av
matematiska avdelningen inom matematisk/teknisk
fysiska sektionsgruppen vid Göteborgs Universitet
och Chalmers Tekniska Högskola framläggs till
offentlig granskning för vinnande av teknisk doktors-
grad lördagen den 1 december 1973 kl 10.15 i sal VG,
Sven Hultins gata 6, Göteborg

UNIV.
LEIDEN
BIBL.

Statistical tests of composite hypotheses and their asymptotic properties

by

Sture Holm

Göteborg 1973

UNIV.
LEIDEN
BIBL.

Printed by

Tryckericentralen

Chalmers University of Technology

Göteborg, October 1973

The dissertation consists of this summary and the following five papers:

(A) Holm S.(1971): On nonparametric asymptotically minimax tests. The Annals of Mathematical Statistics, Vol. 42, No. 2.

(B) Holm S.(1973): The asymptotic minimax character of sequential binomial and sign tests. The Annals of Statistics, Vol. 1, No. 6

(C) Holm S.(1973): Asymptotic power and expected sample size of SPR tests in one-parameter exponential families. Mathematics and Statistics. Essays in Honour of Harald Bergström. Göteborg, Sweden.

(D) Holm S.(1973): Sequential inversion sum tests. To appear in Scandinavian Journal of Statistics.

(E) Holm S.(1973): Asymptotic minimax character of SPR tests in one-parameter exponential classes. To appear in Scandinavian Journal of Statistics.

The papers have appeared in preprint stenciled versions in the series of Technical reports from Department of Mathematics at Chalmers University of Technology and The University of Göteborg under the following numbers: (A) 1970-5, (B) 1972-8, (C) 1973-3, (D) 1973-10, (E) 1973-26.

1. Introduktion, optimality criterias

This thesis and the papers it summarizes are contributions to the theory of asymptotic properties of statistical tests, in particular asymptotic optimality.

There exists a lot of optimality criterias for statistical tests, and at this place it is not possible to review even the main optimum properties that has been shown during the last fifty years. There are e.g. the theories of most powerful tests, most powerful unbiased tests and most powerful invariant tests, the Bayes and minimax properties based on a decision-theoretic approach and minimum expected sample size properties of sequential tests as well as asymptotic optimality in different senses.

Both the choice of optimality criterion and the deduction of optimality is easier for tests of a simple hypothesis against a simple hypothesis than for the cases when one or both of the hypotheses are composite. From a practical point of view however, the last kind of optimality properties are the most valuable ones, and it is desirable to have optimum properties even for tests of composite hypotheses consisting of very large (nonparametric) sets of distributions. It is so because in practice it is almost impossible to state that a distribution is one out of two given, and in many cases there is not even a motivation for stating that a distribution belongs to a parametric class of distributions (i.e. a class where the distributions may be characterized by a simple parameter in a natural way).

One way of obtaining tests of composite hypotheses is to deduce optimal tests with respect to "smaller hypotheses", e.g. with respect to two simple hypotheses in a parametric situation or with respect to a parametric class in a nonparametric situation. This could give an acceptable result, but there remains the problem of

making an over all comparison of the tests obtained for different reduced hypotheses and the question if the best of these tests really have some over all optimum properties.

The first step in structuring a test problem is to choose the two hypotheses, which could be thought of as the sets of distributions for which one or the other of the two possible decisions is preferable. The next step is to order the distributions within each hypothesis with respect to some distance to the other hypothesis. This distance should be interpreted in such a way that for distributions at a longer distance from the opposite hypothesis it is more severe to make the wrong decision. The final step in a decision-theoretic approach is to assign a loss function to distributions and decisions (which naturally depends on the distribution in a hypothesis only through the distance to the opposite hypothesis) and in a sequential problem to fix the cost of experimentation.

Here the first two steps give the rough characterization and the third step a more detailed characterization of the test problem. It is possible to formulate desirable properties for tests and sequences of tests without using the detailed characterization with specified loss and cost. Such properties are e.g. unbiasedness, consistency and minimal maximum probability of wrong decision at a fixed distance from from the opposite hypothesis under some restriction on other probabilities of wrong decision.But the usual decision-theoretic Bayes and minimax optimization requires the detailed knowledge of loss (and in the sequential case of cost) and the solution also depends on the detail behaviour of these functions. This means that there may also be required a considerable amount of computations to get the optimal test for each specific situation.

In practice it is often difficult to specify the loss completely, and it would be valuable to get good tests depending only on some main properties of the loss. One way of doing this is to use asymptotically optimal tests, and in the papers summarized by this thesis two types of asymptotic minimax properties are studied.

For fixed sample size the asymptotic minimax property is defined in A as follows:

Let $R(P;\varphi)$ be the risk of the test with test function φ when the probability measure of the observation is P. Then a sequence $\{\varphi_0^{(N)}: \varphi_0^{(N)} \in I_N\}$ of tests $\varphi_0^{(N)} \in I_N$ is defined to be <u>risk asymptotically minimax</u> in a sequence $\{I_N\}$ of sets I_N of tests when the probability measure belongs to Ω_N if

$$\lim_{N\to\infty} \frac{\sup_{P_N\in\Omega_N} R(P_N;\varphi_0^{(N)})}{\inf_{\{\varphi^{(N)}:\varphi^{(N)}\in I_N\}} \sup_{P_N\in\Omega_N} R(P_N;\varphi^{(N)})} = 1$$

For the sequential case the asymptotic minimax property is defined for a cost of experimentation of c units per trial and c tending to 0. The definition in B reads:

Let P denote the probability measure of the infinite series of trials, and let the risk of a sequential decision rule δ when the cost of experimentation is c units per trial be denoted by $R_c(P;\delta)$. Then a set $\{\delta_0^c : 0<c<\infty\}$ of decision rules δ_0^c is said to be <u>risk asymptotically minimax</u> in a set $\mathcal{D}$ of decision rules and a set $\mathcal{P}$ of probability measures if $\delta_0^c \in \mathcal{D}$ for $0<c<\infty$ and

$$\lim_{c\to 0} \frac{\sup_{p\in\mathcal{P}} R_c(P;\delta_0^c)}{\inf_{\delta\in\mathcal{D}} \sup_{p\in\mathcal{P}} R_c(P;\delta)} = 1$$

The same definition in a specialized form for one-parameter exponential families appears in E.

There can be a weakness in applying asymptotic results as approximations. The convergence can be slow, and since the asymptotic result still holds when changing one member in a sequence, there is really no mening in stating e.g. that a test for N=100 or c=0,001 is a member of an asymptotically minimax set of tests. Naturally however the set of tests 'vary regular' and in our case it is sometimes possible to estimate the quotient of the supremum risk of the used test and the minimax risk. This is done without knowing the minimax test by using parts of the proofs of the asymptotic properties. The loss functions used in the summarized papers are <u>locally linear</u> at the common boundary of the two hypotheses. This means that if d gives a distance parametrization with $d<0$ being the inner of one hypothesis, $d>0$ being the inner of the other hypothesis and $d=0$ being the common boundary, then the loss $L(d;a)$ depends only on the decision a (which equals a_0 or a_1) and the parameter d and

(A) $L(d;a_0)$ and $L(d;a_1)$ are continuous nonnegative functions of d

(B) $L(d;a_1) = 0$ for $d \leq 0$ and $L(d;a_0) = 0$ for $d \geq 0$

(C) $$\lim_{d \to 0} \frac{L(d;a_0)}{d} = \lim_{d \to 0} \frac{L(d;a_1)}{-d} = k \quad \text{where} \quad 0<k< \infty$$

The motivation for using this kind of loss function is that if the 'profits' of making the decisions a_0 and a_1 are continuous decreasing and continuous increasing functions respectively and the 'profits' for d=0 are the same, then the difference between the maximal 'profit' and the 'profit' of making the decisions a_0 or a_1 respectively satisfies conditions (A), (B) and (C). This kind of loss is also discussed by Raiffa and Schlaifer [30] on page 96.

2. Asymptotically minimax tests in the nonparametric two-sample case

The asymptotic minimax properties of tests in the two-sample nonparametric case are studied in A. Let $X_1,\ldots,X_m$ and $Y_1,\ldots,Y_n$ be two independent samples of continuous random variables with component densities f and g and cumulative distribution functions F and G. Further let W be the normalized inversion sum

$$W = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} U_{ij}$$

where $U = \begin{cases} 1 & \text{for } Y_j \geq X_i \\ 0 & \text{for } Y_j < X_i \end{cases}$

and let p denote the probability $p = P(Y_1 \geq X_1) = \int F(u)\, dG(u)$.
In Theorem A.2 is shown that the sequence $\{\varphi_0^{(N)}\}$ of Wilcoxon tests

$$\varphi_0^{(N)} = \begin{cases} 1 & \text{if } W \geq p_0 \\ 0 & \text{if } W < p_0 \end{cases}$$

is asymptotically minimax in the sequence $\{I_N\}$ of sets I_N of all tests when the probability measure belongs to the set Ω_N of all probability measures of $X_1,\ldots, X_m, Y_1,\ldots, Y_n$, if the loss is locally linear in $d = p - p_0$ for some fixed p_0, $0 < p_0 < 1$ and $0 < \lim_{N\to\infty} \frac{m}{N} = \lambda < 1$. Here $\varphi_0^{(N)} = 1$ means taking decision a_1 and $\varphi_0^{(N)} = 0$ taking decision a_0. The notation $N = n + m$ is used.
When applying the theorem A.2 in practice and use the test $\varphi_0^{(m+n)}$ for m and n big we get the approximate power $\Phi\left(\frac{p - p_0}{\sigma_{F,G}}\right)$, where $\sigma_{F,G}$ is the standard deviation of W when the cumulative distributions of X and Y are F and G. The maximal and minimal values of $\sigma_{F,G}$ for fixed p, m and n are given in Lemma A.1 and thus we may get a band of asymptotic powers and asymptotic risks.
For the case $p_0 = \frac{1}{2}$ which is of special interest we have

$$\lim_{N\to\infty} N \sup_{p=\frac{1}{2}} \sigma^2_{F,G} = \frac{1}{4\min(\lambda;1-\lambda)}$$

and

$$\lim_{N\to\infty} \inf_{p=\frac{1}{2}} \sigma^2_{F,G} = \frac{1}{4\max(\lambda;1-\lambda)} - \frac{\min(\lambda;1-\lambda)}{12(\max(\lambda;1-\lambda))^2}$$

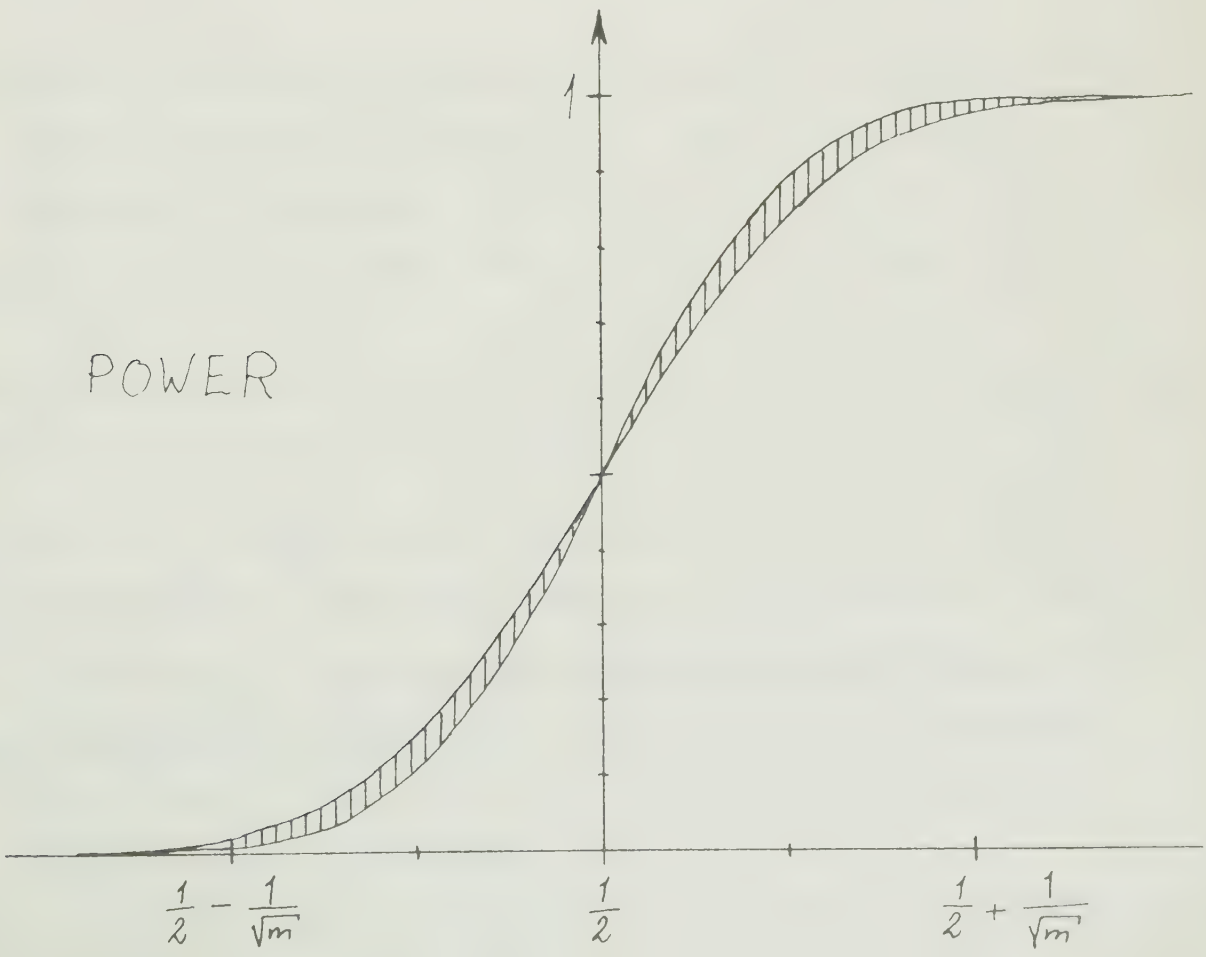
The best asymptotic power and least asymptotic risk is obtained for $\lambda = \frac{1}{2}$, i.e. m=n, in which case

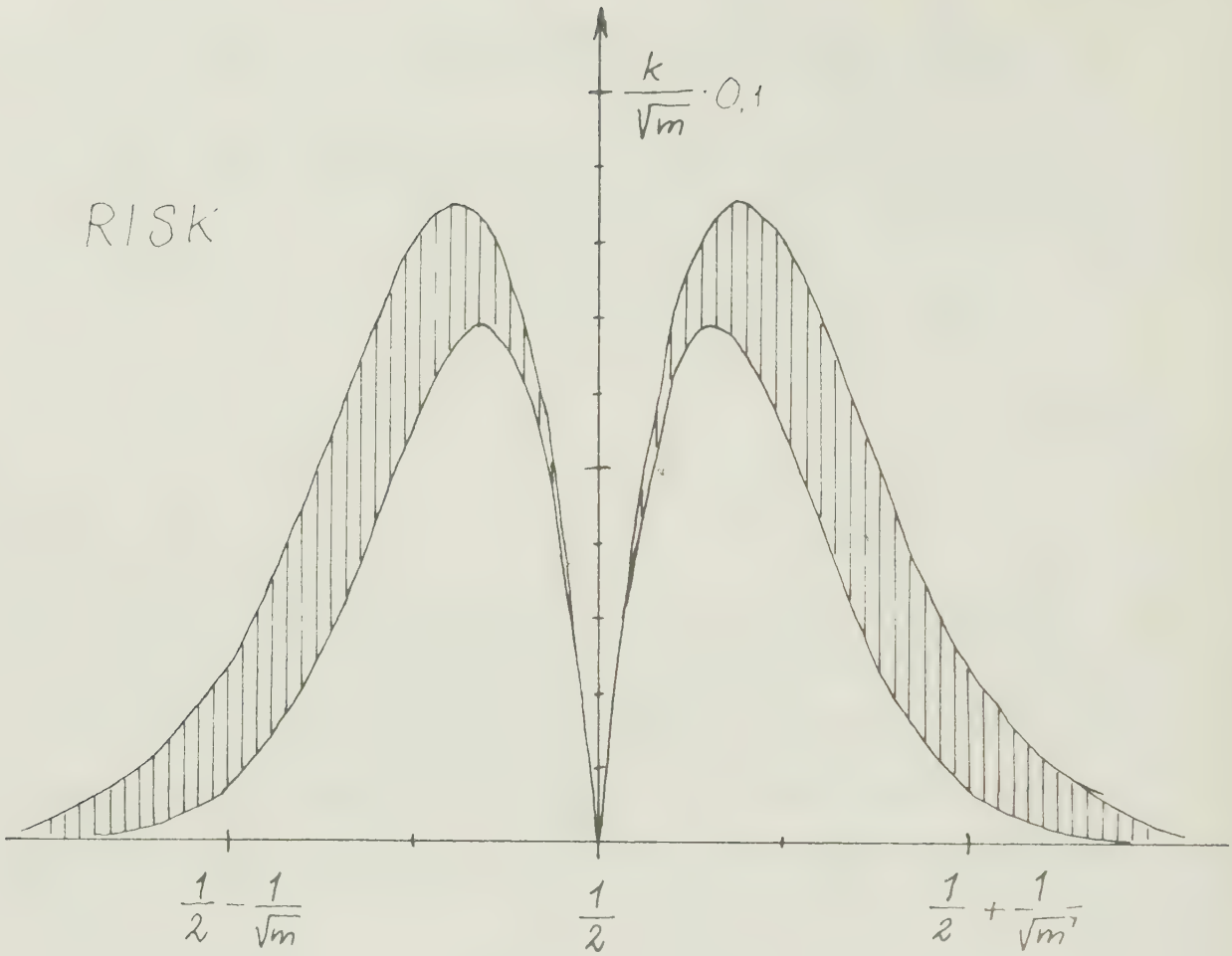
$$\lim_{N\to\infty} \sup_{p=\frac{1}{2}} \sigma^2_{F,G} = \frac{1}{2}$$

and

$$\lim_{N\to\infty} \inf_{p=\frac{1}{2}} \sigma^2_{F,G} = \frac{1}{3}\,.$$

The asymptotic power and risk bands for this case are shown in the following figure. These asymptotic power and risk bands are well suited for planning of experiments and as approximations for the real power and risk bands for big samples. The maximal risk is approximately $0{,}085\,\frac{k}{\sqrt{m}}$ and it occurs for p approximately $\frac{1}{2} \pm \frac{3}{8\sqrt{m}}$.





In the paper A is also shown that the Wilcoxon tests have a power minimax property which is more related to the classical Neyman-Pearson theory. Let $\{\Omega_N(d)\}$ be a partition of Ω_N such that $\omega_N = \Omega_N(0)$ or $\omega_N = \bigcup_{d \leq 0} \Omega_N(d)$ is the hypothesis to be tested and $\{\Omega_N(d): d>0\}$ is a partition of the alternative hypothesis. Further let $\Omega'_N(\Delta) = \bigcup_{d \geq \Delta} \Omega_N(d)$ be the alternatives at a distance of at least Δ from the hypothesis to be tested.
Then a sequence $\{\varphi_0^{(N)}\}$ of level α tests $\varphi_0^{(N)} \in I_N$ is defined to be <u>power asymptotically minimax</u> in a sequence $\{I_N\}$ of sets I_N of level α tests against families $\{\Omega'_N(\Delta): \Delta \in D\}$ of alternative sets for powers $\beta \in B$ if

$$\lim_{N\to\infty} \inf_{P_N \in \Omega'_N(\Delta_N)} E_{P_N}(\varphi_0^{(N)}) =$$

$$= \sup_{\{\varphi^{(N)}:\ \varphi^{(N)} \in I_N\}} \limsup_{N\to\infty} \inf_{P_N \in \Omega'_N(\Delta_N)} E_{P_N}(\varphi^{(N)})$$

for each sequence $\{\Delta_N\}$ satisfying

$$\lim_{N\to\infty} \inf_{P_N \in \Omega'_N(\Delta_N)} E_{P_N}(\varphi_0^{(N)}) = \beta$$

for some $\beta \in B$.

For a sequence of m, n and N=m+n such that $\lim_{N\to\infty} \frac{m}{N} = \lambda$ let the constants a_N be choosen such that the tests

$$\varphi_0^{(N)} = \begin{cases} 1 & \text{if} \quad W > p_0 + \dfrac{a_N}{(\min(m,n))^{1/2}} \\ 0 & \text{otherwise} \end{cases}$$

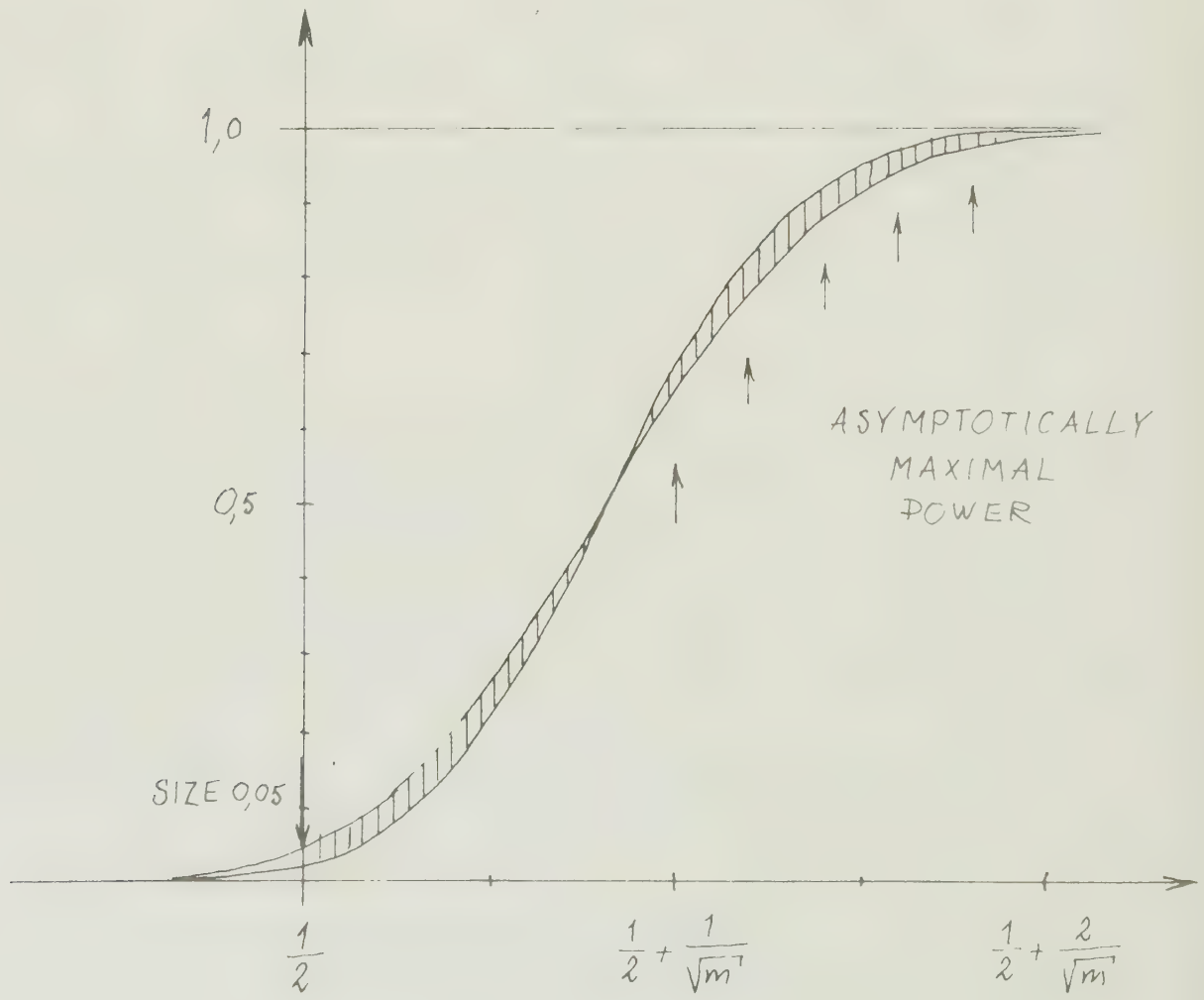
have sizes as close to α as possible for testing the hypothesis $P \in \omega_N = \bigcup_{d\le 0} \Omega_N(d)$, where P is the probability measure of (X,Y) and $\Omega_N(d) = \{P:\ p = p_0+d\}$.

Then it is shown in Theorem A.1 that the sequence $\{\varphi_0^{(N)}\}$ is power asymptotically minimax in the class of all level α tests of $P \in \omega_N$ against the family $\{\Omega'_N(d):\ d>0\}$ for powers $\beta \in (\frac{1}{2};1)$ if $\alpha \le \frac{1}{2}$. The asymptotic power band of these tests is the same as for the risk asymptotically minimax tests for the same p_0 except for a translation. The best asymptotic power is again obtained for $\lambda = \frac{1}{2}$, i.e. m=n, and the figure on next page shows the asymptotic power band for the case $p_0 = \frac{1}{2}$, $\alpha = 0{,}05$ and $\lambda = \frac{1}{2}$

The result in Theorem A.1 may be compared to an earlier result by Doksum [12]. He showed that the sequence of size α Wilcoxon tests of the hypothesis F = G is power asymptotically minimax in a sequence of large sets of nonparametric level α tests of this hypotheses against families $\{\Omega'_N(\Delta):\Delta>0\}$ for powers $\beta \in (\alpha,1]$ with alternative sets $\Omega'_N(\Delta)$ defined by

$$\Omega'_N(\Delta) = \{P_N \in \Omega_N:\ G(x) \le F(x) \text{ for } x \in R \text{ and } \sup_{x\in R} (F(x)-G(x)) \ge \Delta\}.$$

Principally thus his result differs from ours in that he considers only ordered alternatives and shows the minimax property in a slightly reduced class of tests. Further he has to choose Δ_N in such a way that $\lim_{N \to \infty} N^{1/4} \Delta_N = c(\beta) > 0$, i.e. his asymptotic power functions are normalized by $N^{-1/4}$ instead of our $N^{-1/2}$. It is not possible however to say that one result is superior to the other since in different applications it may be natural to introduce one or the other class of alternatives and one or the other distance from the hypotesis to be tested.



3. Asymptotically minimax tests in the one-sample case

The results given in this section are not treated in any of the summarized papers. Using some earlier known results on most powerful tests and the technique of proof used in the paper A, we here deduce asymptotically minimax properties of certain sequences of tests in a nonparametric set of distribution and in one-parameter exponential classes for the one sample case. These results are the fixed sample size analogs of the sequential test results treated in the papers B and E, which will be summarized later.

Let $X_1,\dots, X_N$ be independent random variables with component probability measure P. Further let A be a measurable set in the outcome space of X and $p = P(X \in A)$. Then if the loss is locally linear in $d = p - p_0$ for some p_0, $0 < p_0 < 1$, the sequence $\{ \varphi_0^{(N)} \}$ of tests

$$\varphi_0^{(N)}(x_1,\dots,x_N) = \begin{cases} 1 \text{ if at least } Np_0 \text{ of } x_1,\dots,x_N \text{ belongs to A} \\ 0 \quad \text{othewise} \end{cases}$$

is risk asymptotically minimax in the set of all tests.

In order to prove this we first study the properties of the tests $\varphi_0^{(N)}$. Obviously the distribution of $\varphi_0^{(N)}$ depends only on N and p, and if $\beta(N,p)$ denotes the probability that $\varphi_0^{(N)} = 1$, then by the central limit theorem

$$\lim_{N\to\infty} \beta(N, p_0+\frac{\Delta}{\sqrt{N}}) = \Phi\left(\frac{\Delta}{\sqrt{p_0(1-p_0)}} \right) .$$

From the Berry-Esseen theorem follows that the convergence is uniform for Δ in any interval of the form $[-\Delta_0; \Delta_0]$. Further there exists a constant K such that the loss satisfies

$$L(p;a_0) \le K\,(p-p_0) \quad \text{for} \quad p > p_0$$

and

$$L(p;a_1) \le K\,(p_0-p) \quad \text{for} \quad p < p_0$$

and then by the Chebyshev inequality the risk $R(p; \varphi_0^{(N)})$ of the test $\varphi_0^{(N)}$ satisfies

$$R(p;\varphi_0^{(N)}) \leq \frac{K\,p\,(1-p)}{N\,|p-p_0|} \leq \frac{K}{4\,N\,|p-p_0|}\,.$$

Now if R_0 denotes the maximum of $x\,\Phi(-x)$, we see that the supremum risk of the test equals

$$k\,\frac{R_0}{\sqrt{N}}\,\sqrt{p_0\,(1-p_0)}\;\left(1+\varepsilon(N)\right)$$

where $\lim_{N\to\infty} \varepsilon(N) = 0$, by applying the inequality

$$R(p;\varphi_0^{(N)}) \leq \frac{K}{4\,N\,|p-p_0|}$$

for $|p-p_0| > \dfrac{K}{4\,k\,R_0\sqrt{N}\,\sqrt{p_0(1-p_0)}}$, and the local linearity and the uniform convergence for $|p-p_0| < \dfrac{K}{4\,k\,R_0\sqrt{N}\,\sqrt{p_0(1-p_0)}}$.

If x_0 denotes the solution of $\Phi(-x) - x\,\varphi(x) = 0$, i.e. the argument maximizing $x\,\Phi(-x)$ we now study the alternatives

$$p'_N = p_0 - \sqrt{p_0(1-p_0)}\,\frac{x_0}{\sqrt{N}}$$

and

$$p''_N = p_0 + \sqrt{p_0(1-p_0)}\,\frac{x_0}{\sqrt{N}}\,.$$

Let P_0 be any probability measure such that $P_0(A) = p_0$ and define probability measures P'_N and P''_N by

$$P'_N(B) = \frac{p'_N}{p_0}\,P_0(B\cap A) + \frac{1-p'_N}{1-p_0}\,P_0(B\smallsetminus A)$$

and

$$P''_N(B) = \frac{p''_N}{p_0}\,P_0(B\cap A) + \frac{1-p''_N}{1-p_0}\,P(B\smallsetminus A).$$

The minimax test with respect to these two alternatives only is given by the Neyman-Pearson lemma, and since the Radon-Nikodym derivative of P''_N with respect to P'_N is

$$\frac{p''_N}{p'_N}\,I_A + \frac{1-p''_N}{1-p'_N}\,(1-I_A)$$

where I_A is the indicator of A, the minimax test is a sign test for the set A. But the monotonity of the powers as functions of the

rejection limit and the earlier discussions now show that the minimax risk can not be smaller than

$$k \frac{R_0}{\sqrt{N}} \sqrt{p_0(1-p_0)} \; (\; 1 - \; \varepsilon(N) \;)$$

where $\lim_{N \to \infty} \varepsilon(N) = 0$.

This proves the stated result that the proposed sign test is asymptotically minimax.

Next we are going to study one-parameter exponential classes. We use the natural parametrization, i.e. consider probability measures with a density of the form

$$C(\Theta) \; e^{\Theta T(x)}$$

with respect to some measure $\mu(x)$. See the papers C and E for discussion of this parametrization and basic properties. The hypotheses are supposed to be of the forms $\Theta \leq \Theta_0$ and $\Theta > \Theta_0$ for some Θ_0, which is an inner point of the natural parameter interval, and the decision is to be based on a set $X_1, \ldots, X_N$ of independent random variables.

Then the set $\{ \varphi_0^{(N)} \}$ of tests

$$\varphi_0^{(N)}(x_1, \ldots, x_N) = \begin{cases} 1 \text{ if } \sum_{k=1}^{N} T(X_k) + N \dfrac{C'(\Theta_0)}{C(\Theta_0)} \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

is asymptotically minimax in the set of all tests if the loss is locally linear in $d = \Theta - \Theta_0$ and the loss $L(\Theta; a)$ also satisfies the condition:

$$L(\Theta; a_0) \frac{\sigma^2(\Theta)}{(\mu(\Theta) - \mu(\Theta_0))^2} \text{ is bounded for } \Theta > \Theta_0 + \gamma \text{ for}$$

any fixed $\gamma > 0$, and

$$L(\Theta; a_1) \frac{\sigma^2(\Theta)}{(\mu(\Theta) - \mu(\Theta_0))^2} \text{ is bounded for } \Theta < \Theta_0 - \gamma \text{ for}$$

any fixed $\gamma > 0$,

where $\mu(\Theta) = E_\Theta(T(X))$ and $\sigma^2(\Theta) = \mathrm{Var}_\Theta(T(X))$.

The proof follows the same lines as the previous one. The random variable $\sum_{k=1}^{N} (T(X_k) + \frac{C'(\Theta_0)}{C(\Theta_0)})$ is asymptotically normal, and if the power of the test for sample size N and parameter Θ is denoted by $\beta(N;\Theta)$ then

$$\lim_{N\to\infty} \beta(N;\Theta_0 + \frac{\Delta}{\sqrt{N}}) = \Phi(\frac{\Delta\,\mu'(\Theta_0)}{\sigma(\Theta_0)}) = \Phi(\Delta\sigma(\Theta_0)).$$

The convergence is uniform for Δ in any interval of the form $[-\Delta_0;\Delta_0]$.

By the Chebyshev inequality the risk $R(\Theta;\varphi_0^{(N)})$ is bounded by

$$L(\Theta;a_0)\,\frac{\sigma^2(\Theta)}{N(\mu(\Theta)-\mu(\Theta_0))^2} \quad \text{for} \quad \Theta > \Theta_0$$

and by

$$L(\Theta;a_1)\,\frac{\sigma^2(\Theta)}{N(\mu(\Theta)-\mu(\Theta_0))^2} \quad \text{for} \quad \Theta < \Theta_0.$$

From the imposed condition and the local linearity now follows the existance of a constant Δ_0 such that

$$R(\Theta_0 + \frac{\Delta}{\sqrt{N}};\varphi_0^{(N)}) \le k\,\frac{R_0}{2\,\sigma(\Theta_0)\sqrt{N}} \quad \text{for} \quad |\Delta| \ge \Delta_0,$$

and the uniformity in the convergence implies that the maximal risk of the test $\varphi_0^{(N)}$ is bounded by

$$k\,\frac{R_0}{\sigma(\Theta_0)\sqrt{N}}\,(1+\varepsilon(N))$$

where $\lim_{N\to\infty} \varepsilon(N) = 0$.

The minimax test with respect only to $\Theta' = \Theta_0 - \frac{1}{\sigma(\Theta_0)\sqrt{N}}$ and $\Theta'' = \Theta_0 + \frac{1}{\sigma(\Theta_0)\sqrt{N}}$ must (by the Neyman-Pearson lemma) be based on the random variable $\sum_{k=1}^{N} T(X_k)$, and thus the minimax

risk can not be smaller than

$$k \frac{R_0}{\sigma(\Theta_0)\sqrt{N}} (1 - \varepsilon(N))$$

where $\lim_{N \to \infty} \varepsilon(N) = 0$.

This proves the stated asymptotically minimax property.

The practical implications of the two results given in this section are that for big samples the proposed test are 'almost minimax'. Observe that they are determined only by the 'contact points' p_0 or Θ_0 between the hypotheses, but yet they have good decision-theoretic properties. An upper bound of the relative difference between the obtained risk and the minimax risk is $2\varepsilon(N)$, which could be estimated in special cases by the Berry-Esseen theorem and the non-linearity of the loss function.

In the non-parametric situation treated in section 2 and equal sample sizes $m = n$ it is now possible to get two quite different sequences of asymptotically minimax tests. Besides a sequence of Wilcoxon tests being asymptotically minimax, a sequence of sign tests for the event $Y>X$ also is. In this case we namely have n=m trials of the same kind (whose results are the two-dimensional random variables $(X_k;Y_k)$ $k=1,2,\ldots,n$) and the loss is locally linear in the probability of the event $Y>X$ in these trials. This may seem surprising, but it is to be noted that under asymptotically minimax alternatives the Wilcoxon test acts as a sign test, and that the difference between the tests in concentrated to a null set under these alternatives. For other distributions however this set may have a great probability and the power and risk for the two tests may differ considerably.

It is also possible to get a third quite different sequence of tests, which is asymptotically minimax. The test problem is invariant under

the finite group consisting in permuting the Y:s (or the X:s). The Wilcoxon test is invariant under this group, but the sign test is not, and thus we get a test which has a risk smaller than or equal to the risk of the sign test by averaging over the group. This averaged sign test, which is a randomized test, is then also asymptotically minimax. It may be observed that a sign test and the corresponding averaged sign test are minimax for each fixed sample size m=n too. The risk of a sign test is a continuous function of p, monotonically decreasing in the limit in the sign test for $p>p_0$ and monotonically increasing in this limit for $p<p_0$. In order to get the minimax test we first adjust the limit to get equal maximum risks for $p>p_0$ and $p<p_0$ (and randomize if necessary). Studying next two alternatives of the kind used in proving asymptotic minimaxity with p-values giving maximal risk for $p>p_0$ and $p<p_0$, we see that the maximum risk can not be decreased. The averaged sign test will have the same maximum risk. These minimax tests thus require some computations for each special case, and for big samples there is little advantage in using them compared to the asymptotically minimax tests.

4. Asymptotically minimax sequential tests for probabilities

Contributions to the theory of sequential tests for sampling from a single distribution has been made in the papers B, C and E. Chronologically it was first shown (in the paper B) that a certain set of sequential tests is asymptotically minimax for a two-point (binomial) distribution and that an analogous set of tests is asymptotically minimax when the loss depends only on the probability of an event in an arbitrary repeted trial. Later it was discovered (the paper C) that an unified form of the asymptotic power and expected sample size of this kind of tests existed in one-parameter exponential classes. And finally it was proved (in the paper E) that this kind of tests is asymptotically minimax for one-parameter exponential classes with mild restrictions imposed.

Although the binomial case now could be treated as a special case of exponential class, we prefer to handle first the binomial case, and later handle the extension to general exponential classes, because this is the way it is done in the papers.

Let $X_1, X_2, \ldots$ be independent random variables and suppose that $P(X_k=0) = 1-p$ and $P(X_k=1) = p$, where $0<p<1$, and let δ_0^c be the sequential decision rule, that continues sampling at stage n if

$$\left| \sum_{k=1}^{n} (X_k-p_0) \right| < \varrho_c ,$$ takes decision a_0 if $$\sum_{k=1}^{n} (X_k-p_0) < -\varrho_c$$ and takes decision a_1 if $$\sum_{k=1}^{n} (X_k-p_0) > \varrho_c ,$$ where

$$\varrho_c = \frac{1}{2} k^{1/3} p_0^{2/3} (1-p_0)^{2/3} \eta \, c^{-1/3}$$

$$\eta = \left(v^3 / (v+\sinh v) \right)^{1/3}$$

$$(1 + e^{-v}) (v + \sinh v) = 2 v \sinh v.$$

(The solution of the third equation is unique.)

Then it is shown in Theorem B.4 that the set $\{\delta_0^c : 0 < c < \infty\}$ is asymptotically minimax in the set of all decision rules, if the loss is locally linear in $d = p - p_0$.

Further let $P_p^c(a_0)$ denote the probability that the final decision with the rule δ_0^c will be a_0, $E_p^c(N)$ denote the expected sample size of the rule δ_0^c and $R_c(p;\delta_0^c)$ denote the risk of the rule δ_0^c when the parameter is p.

Then it is shown in the Theorems B.1, B.2 and B.3 that

$$\lim_{c \to 0} P_{p_0+\Delta_c}^c(a_0) = \frac{1}{e^{m\eta}+1},$$

$$\lim_{c \to 0} c^{2/3} E_{p_0+\Delta_c}^c(N) = \frac{(e^{m\eta}-1)\,k\,\vartheta^{-1}}{2m(e^{m\eta}+1)}$$

and

$$\lim_{c \to 0} c^{-1/3} R_c(p_0+\Delta_c;\delta_0^c) =$$

$$= k\,\vartheta^{-1}\left\{\frac{|m|}{e^{|m|\eta}+1} + \frac{\eta(e^{m\eta}-1)}{2m(e^{m\eta}+1)}\right\}$$

where

$$\Delta_c = c^{1/3} m \vartheta^{-1}$$

$$\vartheta = p_0^{-1/3}(1-p_0)^{-1/3} k^{1/3}$$

k is the constant in the condition of local linearity.

All three convergencies are uniform for m in any interval of the form $[-m_0;m_0]$.

These four theorems are proved by help of five lemmas. Lemma B.1 gives the behaviour of the solution (in t) of $E(e^{(X-p_0)t}) = 1$ for $p \approx p_0$, which is an equation highly connected with the classical Wald technique of obtaining properties for SPR tests. Lemma B.2 gives estimates for $P_p^c(a_0)$ and Lemma B.3 gives specialized versions of these estimates useful in the tails. In lemma B.4 are given upper and lower estimates of the expected sample size. These estimates are not good for p near p_0 however, and in Lemma B.5 finally is given

the inequality

$$\left| \frac{d}{dp} E_p^c(N) \right| \leq \frac{1}{p(1-p)} \{ (\varrho_c + 1) E_p^c(N) + |p-p_0| E_p^c(N^2) \}$$

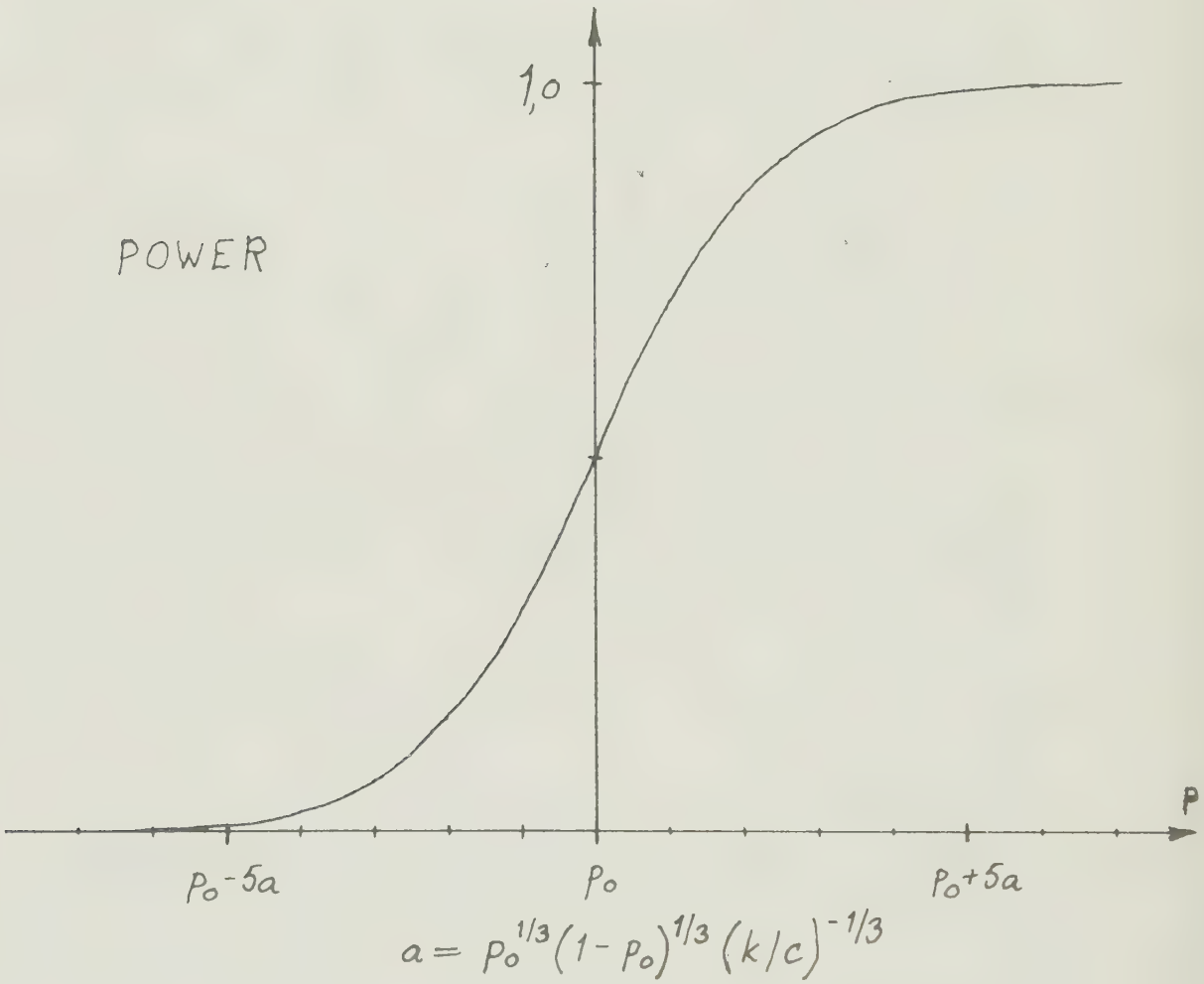
which is good enough to handle the problems for p in a neighbouhood of p_0. Lemmas B.1-B.4 uses the standard technique in SPR analysis in their proof, while Lemma B.5 and the use of that kind of estimate seems to be new. This kind of estimate will be still more important in the general exponential class case, where the classical technique will be completely skipped.

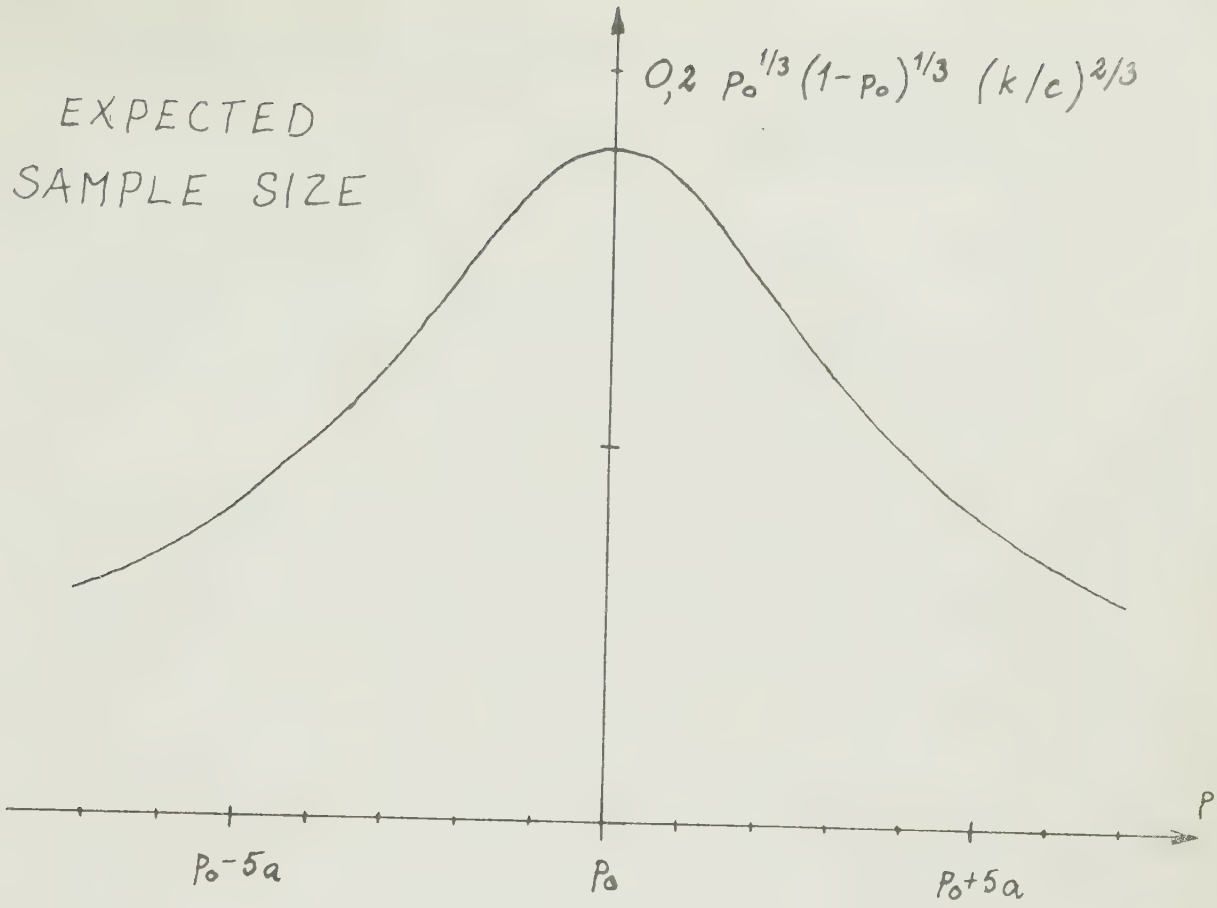
The way of proving the theorems is now first to use Lemmas B.1, B.2, B.4 and B.5 to prove convergence and uniformity in Theorems B.1 and B.2. From these is obtained the convergence and uniformity in Theorem B.3. This theorem together with the tail estimates in Lemma B.3 is used to get an upper bound of the maximum risk of the proposed test. Finally is studied two specially chosen alternatives and a lower bound of the minimax risk for these alternatives only is obtained and it is shown that the quotient of these bounds tends to 1, which proves the asymptotic minimax property stated in Theorem B.4.

The result in Theorem B.4 is also extended to a nonparametric formulation, where the sequentially performed trials may be of any kind and the loss depende only on the probability p of a fixed event A in these trials. Then if the loss is locally linear in $d = p - p_0$ for some p_0, $0 < p_0 < 1$, the test obtained from the previous case by putting X_k equal to the indicator of the event A in trial number k is asymptotically minimax in the set of all tests. Obviously these tests have the same properties (depending only on c and p) as the tests used in the binomial case, and it is also possible to find a pair of distributions for which the risk is the same as the risk for the pair of distributions used in the bino-

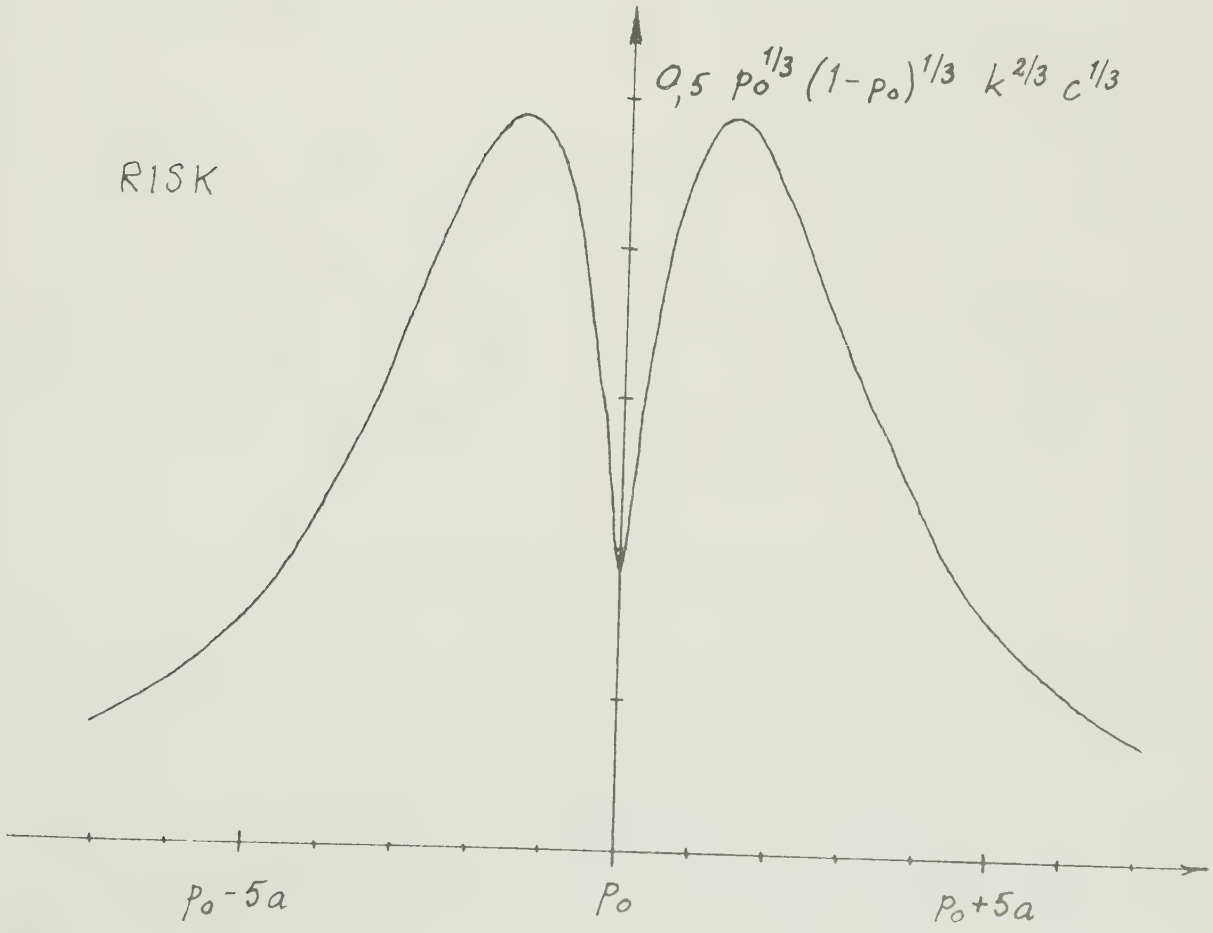
mial case.

The asymptotically approximate power, expected sample size and risk for the proposed tests are given in the following figures.





$$a = p_o^{1/3}(1-p_o)^{1/3}(k/c)^{-1/3}$$



$$a = p_o^{1/3}(1-p_o)^{1/3}(k/c)^{-1/3}$$

2. Asymptotically minimax sequential tests in one-parameter exponential classes

Suppose that $X_1, X_2, \ldots$ are independent random variables with the same distribution belonging to an one-parameter exponential class with a natural parametrization, i.e. having a density of the form

$$C(\Theta)\ e^{\Theta T(x)}$$

with respect to some measure $\mu(x)$. Suppose further that the hypothesis to be tested is $\Theta \leq \Theta_0$ for some Θ_0 which is an inner point of the natural parameter interval and that the alternative hypothesis is $\Theta > \Theta_0$.

A quite natural set of tests in this case are the tests that continue sampling as long as

$$b < \sum_{k=1}^{n} \left(T(X_k) + \frac{C'(\Theta_0)}{C(\Theta_0)}\right) < a$$

and make the statements $\Theta \leq \Theta_0$ or $\Theta > \Theta_0$ depending on which limit is first reached.

In the paper C are given asymptotic properties of these tests when a and b tends to ∞ and $-\infty$ at proportional rates. If $p_\Theta(\alpha,\beta,M)$ denotes the probability of making the final statement $\Theta > \Theta_0$ when the parameter is Θ and the limits are $a = \alpha M$ and $b = \beta M$, then it is shown in Theorem C.1 that under mild restrictions on the tails of the distribution we have

$$\lim_{M\to\infty} p_{\Theta_0+\frac{\Delta}{M}}(\alpha,\beta,M) = \frac{e^{-2\beta\Delta} - 1}{e^{-2\beta\Delta} - e^{-2\alpha\Delta}} \quad \text{for} \quad \Delta \neq 0.$$

And if $e_\Theta(\alpha,\beta,M)$ denotes the sample size of the same test, then it is shown in Theorem C.2 that under the same restrictions

$$\lim_{M\to\infty} \frac{1}{M^2} e_{\Theta_0+\frac{\Delta}{M}}(\alpha,\beta,M) =$$

$$= \frac{1}{\Delta \operatorname{Var}_{\Theta_0} T(X)} \cdot \frac{\alpha\,(e^{-2\beta\Delta} - 1) + \beta\,(1 - e^{-2\alpha\Delta})}{e^{-2\beta\Delta} - e^{-2\alpha\Delta}}$$

for $\Delta \neq 0$.

In both cases classical Wald methods are used.

These results are reproved in the paper E, where weak convergence technique is used and the restrictions may be dropped. There is also proved that the convergences are uniform in any interval of the form $[-\Delta_0; \Delta_0]$. This is done by showing equicontinuity of $p_{\Theta_0 + \frac{\Delta}{M}}(\alpha,\beta,M)$ and $\frac{1}{M^2} e_{\Theta_0 + \frac{\Delta}{M}}(\alpha,\beta,M)$ based on two inequalities stated in Lemma E.1. If $Z_n = \sum_{k=1}^{n} T(X_k) + n \frac{C'(\Theta_0)}{C(\Theta_0)}$, N is the random sample size obtained in a test with limits a and b, P_Θ denotes the probability when the parameter is Θ and E_Θ denotes the expectation in this case the inequalities reads

$$\left| \frac{d}{d\Theta} E_\Theta(N) \right| \leq E_\Theta(N\,|Z_N|) + E_\Theta(N^2)\,|\mu(\Theta) - \mu(\Theta_0)|$$

and

$$\left| \frac{d}{d\Theta} P_\Theta(Z_N \geq a) \right| \leq \frac{1}{2} E_\Theta(N)\,|\mu(\Theta) - \mu(\Theta_0)| + \frac{1}{2} E_\Theta(|Z_N|)$$

where $\mu(\Theta) = E_\Theta(T(X))$.

These inequalities which reminds of the inequality in Lemma B.5 requires the condition (denoted by (D) in the paper E):

$$\gamma_1(\Theta) = \inf_{r>0} E_\Theta(T(X)+r \mid T(X) \leq -r < 0)$$

and

$$\gamma_2(\Theta) = \sup_{r>0} E_\Theta(T(X)-r \mid T(X) \geq r > 0)$$

are bounded in a neighbourhood of each interior point of the natural parameter interval.

Now suppose that the loss is locally linear in $d = \Theta - \Theta_0$ and let δ_0^c be the test with limits $a = -b = \frac{1}{2}(\operatorname{Var}_{\Theta_0}(T(X)))^{2/3} k^{1/3} \eta\, c^{-1/3}$

where k is the constant in the condition of local linearity and η is determined by

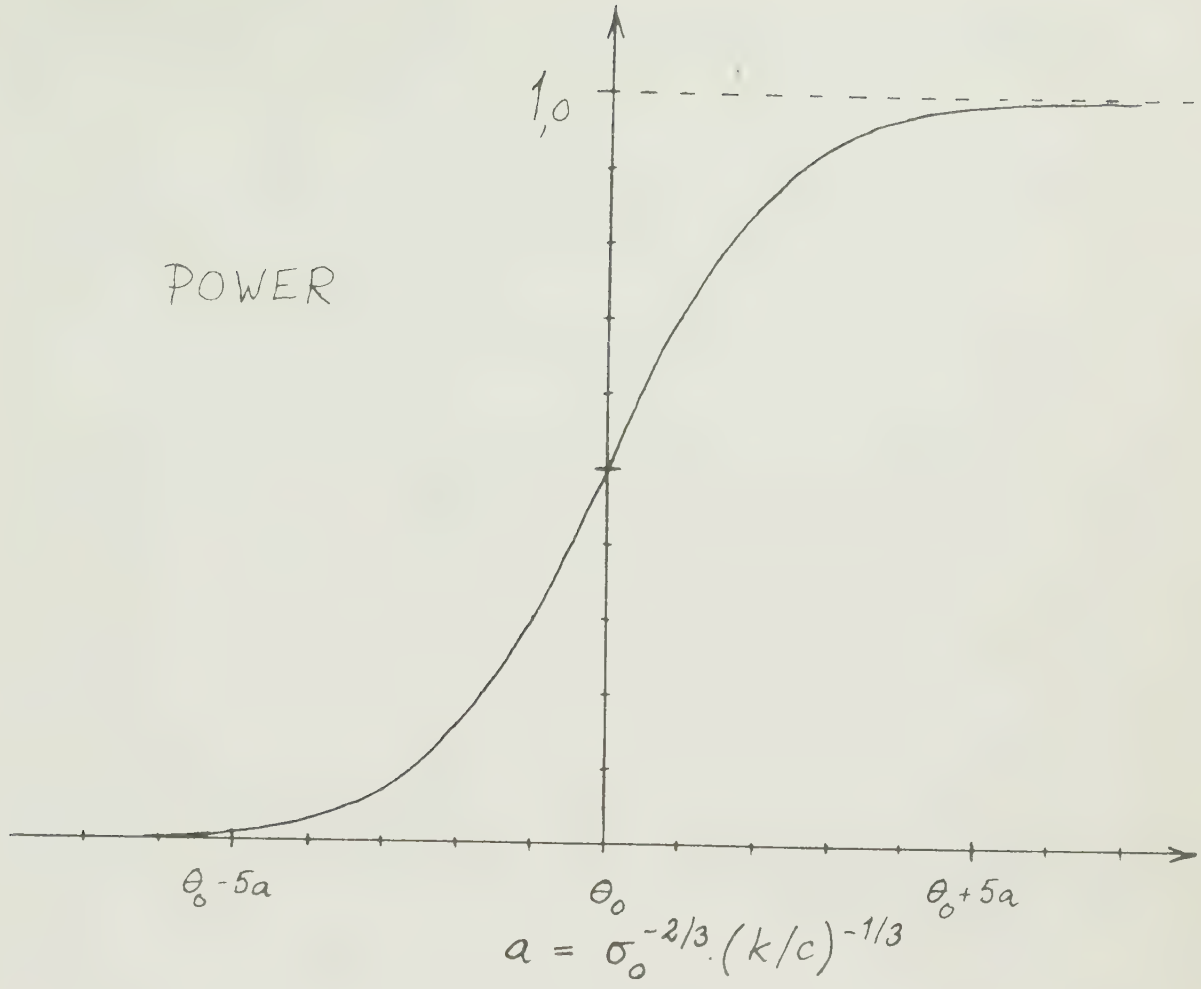
$$\eta = (v^3/(v + \sinh v))^{1/3}$$

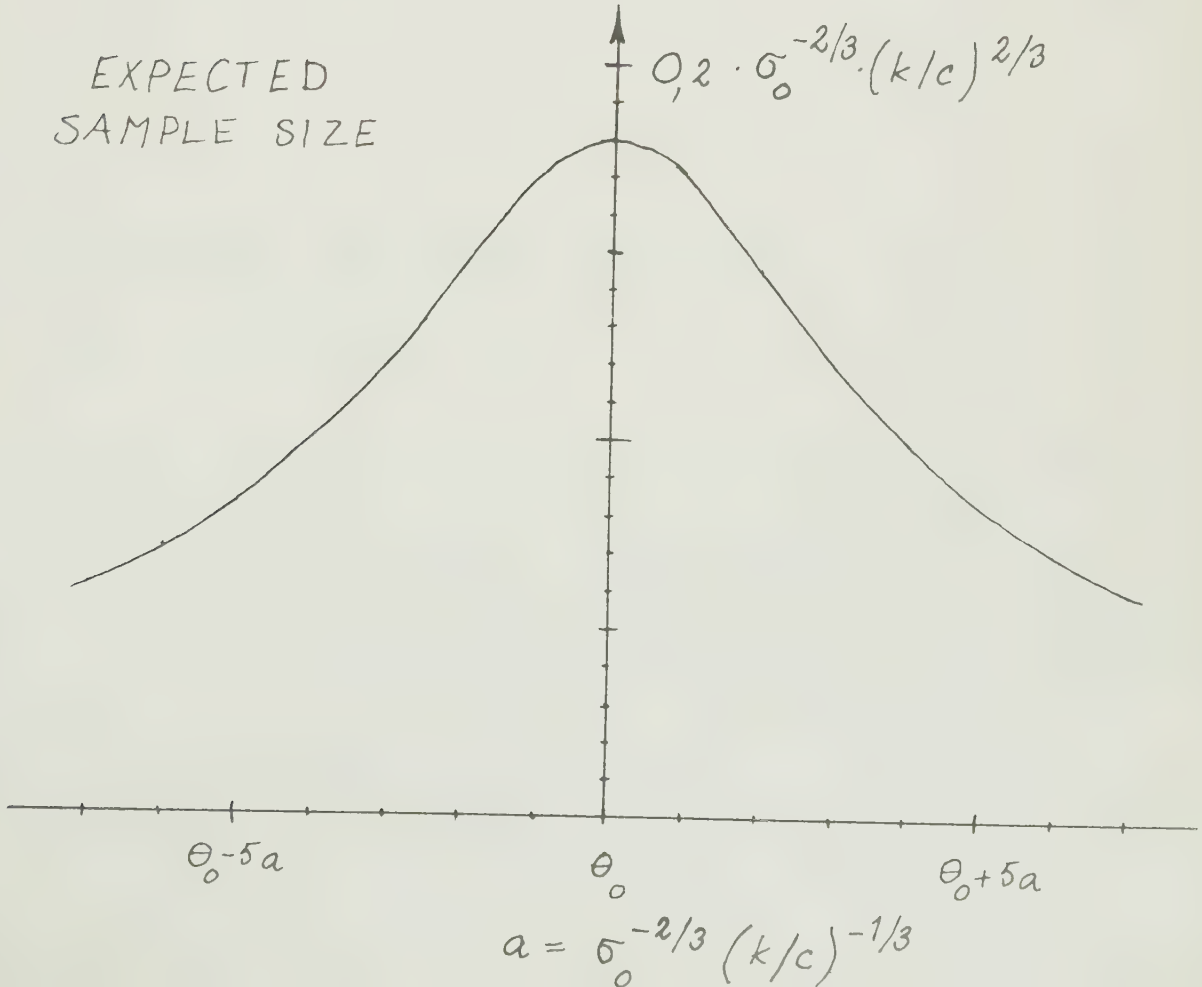
$$(1 + e^v) (v + \sinh v) = 2 v \sinh v.$$

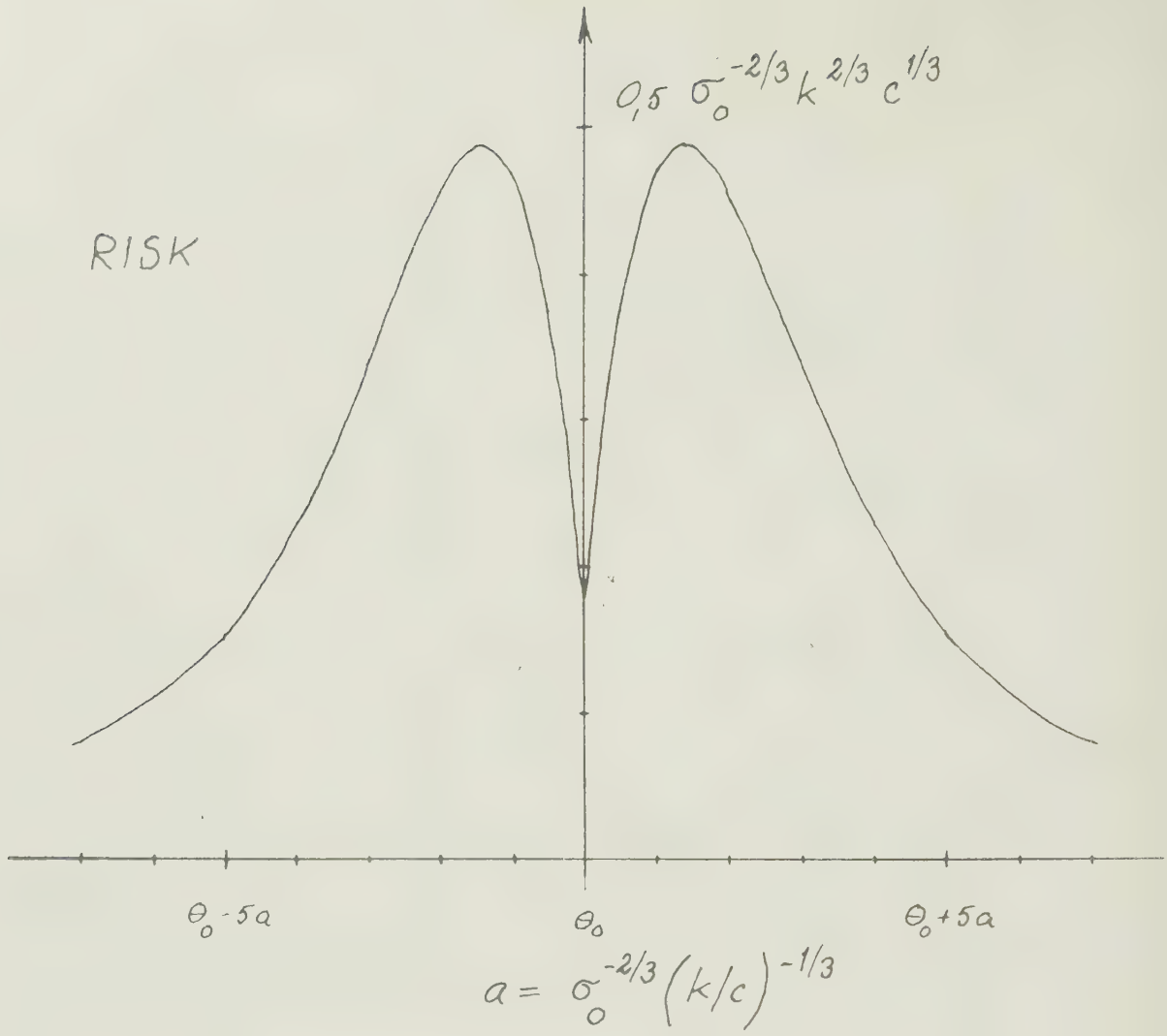
Then it is shown in Theorem E.2 that the set $\{ \delta_0^c : 0 < c < \infty \}$ is asymptotically minimax in the set of all tests if the exponential class satisfies the earlier quoted condition (D) and another mild condition (F) on conditional expectation of T(X) together with a condition (E) that the loss will not grow too fast for distant alternatives.

The practical implication of the theorem is that for small cost c per trial the test δ_0^c may be used as an 'almost minimax' test. This test is determined by the 'contact point' Θ_0 between the hypotheses and the quotient k/c of the local linearity constant k and cost constant c.

Approximate power, expected sample size and risk obtained from the asymptotic properties are shown in the following three figures. The constant σ_0 which appears at some places in the figures stands for $(\mathrm{Var}_{\Theta_0} (T(X)))^{1/2}$.


1,0
POWER
$\theta_o - 5a$
θ_o
$\theta_o + 5a$
$a = \sigma_o^{-2/3} \cdot (k/c)^{-1/3}$


EXPECTED
SAMPLE SIZE
$0,2 \cdot \sigma_o^{-2/3} \cdot (k/c)^{2/3}$
$\theta_o - 5a$
θ_o
$\theta_o + 5a$
$a = \sigma_o^{-2/3} (k/c)^{-1/3}$


RISK
$0{,}5\ \sigma_0^{-2/3} k^{2/3} c^{1/3}$
$\theta_0 - 5a$
θ_0
$\theta_0 + 5a$
$a = \sigma_0^{-2/3} (k/c)^{-1/3}$

6. Sequential tests for sampling from two distributions belonging to a nonparametric class

Let X and Y be independent continuous random variables with cumulative distribution functions F and G respectively, put $p = P(X < Y) = \int_{-\infty}^{\infty} F(x)\, dG(x)$ and suppose we want to test the hypothesis $p \leq p_0$ against $p > p_0$ as in section 2. But now we are interested in sequential tests, and since we know from section 2 that equal sample sizes give the best asymptotic properties for fixed sample sizes, we now treat only the case when (X_1,Y_1) is first observed, (X_2,Y_2) is next observed and so on, where all variables $X_1, Y_1, X_2, Y_2, \ldots$ are independent, the X:s having cumulative distribution function F and the Y:s having the cumulative distribution function G.

The kinds of tests treated in the paper D are based on the inversion sum

$$U(n) = \sum_{i=1}^{n} \sum_{j=1}^{n} U_{ij}$$

where

$$U_{ij} = \begin{cases} 1 & \text{if } Y_j > X_i \\ 0 & \text{if } Y_j \leq X_i. \end{cases}$$

A sequential test which continues sampling as long as

$$b < \frac{U(n) - n^2 p_0}{2n} < a,$$

accepts the hypothesis $p \leq p_0$ if the lower limit is reached and rejects the hypothesis $p \leq p_0$ if the upper limit is reached is called a (non-truncated) sequential inversion sum test with parameters (p_0,a,b). And a truncated sequential inversion sum test with parameters (M,p_0,a,b) is a test of the same kind, except that if M pairs of random variables are observed, no further experimentation is done, the hypothesis $p \leq p_0$ is accepted if $(U(M) - M^2 p_0)/2M \leq 0$

and rejected if $(U(M) - M^2 p_0)/ 2M > 0$.

In contrast to the usual SPR tests all earlier observations must be available at each step in the procedure. The calculations needed ar however very simple since

$$U(n) = U(n-1) + \sum_{j=1}^{n-1} U_{nj} + \sum_{i=1}^{n} U_{in}$$

which means that we have to add the inversions among X_n and the 'old' $Y_1,\ldots, Y_{n-1}$ and the inversions among Y_n and all $X_1,\ldots., X_n$ to the previous inversion sum. This can easily be done by using labels of different colours for the two kinds of observations and arrange these in increasing order.

Some properties of sequential inversion sum tests are deduced in the paper D. In Theorem D.1 is proved that the stopping time is finite with probability one and the moments of all orders of the stopping time are finite for any continuous distributions of the X- and Y-variables. In the proof the Lemma D.1 on convergence of relative frequencies plays an important role.

A great deal of the paper D is devoted to weak convergence of a rank process associated with the sequential inversion sum test to a Brownian motion. In four lemmas (D.2-D.5) are proved different inequalitities which are used to prove the tightness in Theorem D.1 and convergence in Theorems D.2 and D.3. The main idea in the proof is to use an additive process

$$Z(n) = \frac{1}{2} \sum_{k=1}^{n} (F(Y_k) + 1 - G(X_k) - 2p)$$

as an approximation of

$$S(n) = \frac{U(n) - n^2 p}{2n} .$$

Once this result is obtained it is possible to get asymptotic power and expected sample size for the sequential inversion sum test.In the

convergence to the asymptotic power given in Theorem D.5 the test limits a and b has to tend to ∞ and $-\infty$ at rates which are inverse to the rate at which the p-value of the distributions tend to p_0. The limiting power for the nontruncated case is

$$\frac{e^{-2\mu\beta/\sigma} - 1}{e^{-2\mu\beta/\sigma} - e^{-2\mu\alpha/\sigma}} \quad \text{for} \quad \mu \neq 0 \quad \text{and}$$

$$\frac{-\beta}{\alpha - \beta} \quad \text{for} \quad \mu = 0 \,,$$

where the sequence of cumulative distribution functions F_n and G_n are chosen such that

$$\lim_{n\to\infty} \sigma^2_{F_n,G_n} = \sigma^2 \quad \text{and} \quad \lim_{n\to\infty} \sqrt{n}\,(\mu_{F_n,G_n} - p_0) = 2\mu\sigma \,,$$

with μ_{F_n,G_n} and $\sigma^2_{F_n,G_n}$ being the expectation and variance of $\frac{1}{2}(F_n(Y) + 1 - G_n(X))$ when the cumulative distribution functions of X and Y are F_n and G_n. The limits in the tests are then chosen to be $a = \alpha\sqrt{n}$ and $b = \beta\sqrt{n}$.

The asymptotic expected sample size normalized by a factor $\frac{1}{n}$ is given in Theorem D.6. Under the same conditions as for the asymptotic power we have the limit

$$\frac{\frac{\alpha}{\sigma}(e^{-2\beta\mu/\sigma} - 1) + \frac{\beta}{\sigma}(1 - e^{-2\alpha\mu/\sigma})}{(e^{-2\beta\mu/\sigma} - e^{-2\alpha\mu/\sigma})} \quad \text{for} \quad \mu \neq 0$$

It has not been possible to show that some sequence of sequential inversion sum tests is asymptotically minimax for losses that are locally linear in $d = p - p_0$, but in section 7 of the paper D is shown that a certain sequence of tests has the weaker optimum property that no sequence of tests can have a smaller maximum of the asymptotic risk for $\mu \neq 0$ if $0{,}16 \leq p_0 \leq 0{,}84$. This is done by studying the asymptotic risk band of the tests, finding a sequence

of pairs of alternatives giving maximum asymptotic risk and show that the maximal asymptotic risk of these can not be decreased. The reason why the property holds only for p_0 in a subinterval of (0;1) and not for the whole interval is that the character of the maximum poit is changed for p_0 near 0 and 1.

Since the loss here is locally linear in the probability $p = P(Y>X)$ we could use the theory developed in section 4 to say that a sequential sign test of the kind given there (counting events $X_1<Y_1$, $X_2<Y_2$,...) is asymptotically minimax. The asymptotic risk of this test has the same maximum as the maximum of the asymptotic risk band of the sequential inversion sum test, but for distributions not giving maximum, the inversion sum test in most cases gives a lower risk (at least for high risks). Compare to section 4.

In the following figures are shown the asymptotic power band, the asymptotic expected sample size band and the asymptotic risk band of the sequential inversion sum test for the case $p_0 = \frac{1}{2}$, which is of special practical interest.

The asymptotically approximate power is

$$\frac{1}{1 + e^{m \eta \sigma^2_{max}/\sigma^2}}$$

the asymptotically approximate expected sample size is

$$4^{-1/3} (k/c)^{2/3} \frac{\eta}{2m} \frac{e^{m \eta \sigma^2_{max}/\sigma^2} - 1}{e^{m \eta \sigma^2_{max}/\sigma^2} + 1}$$

and the asymptotically approximate risk is

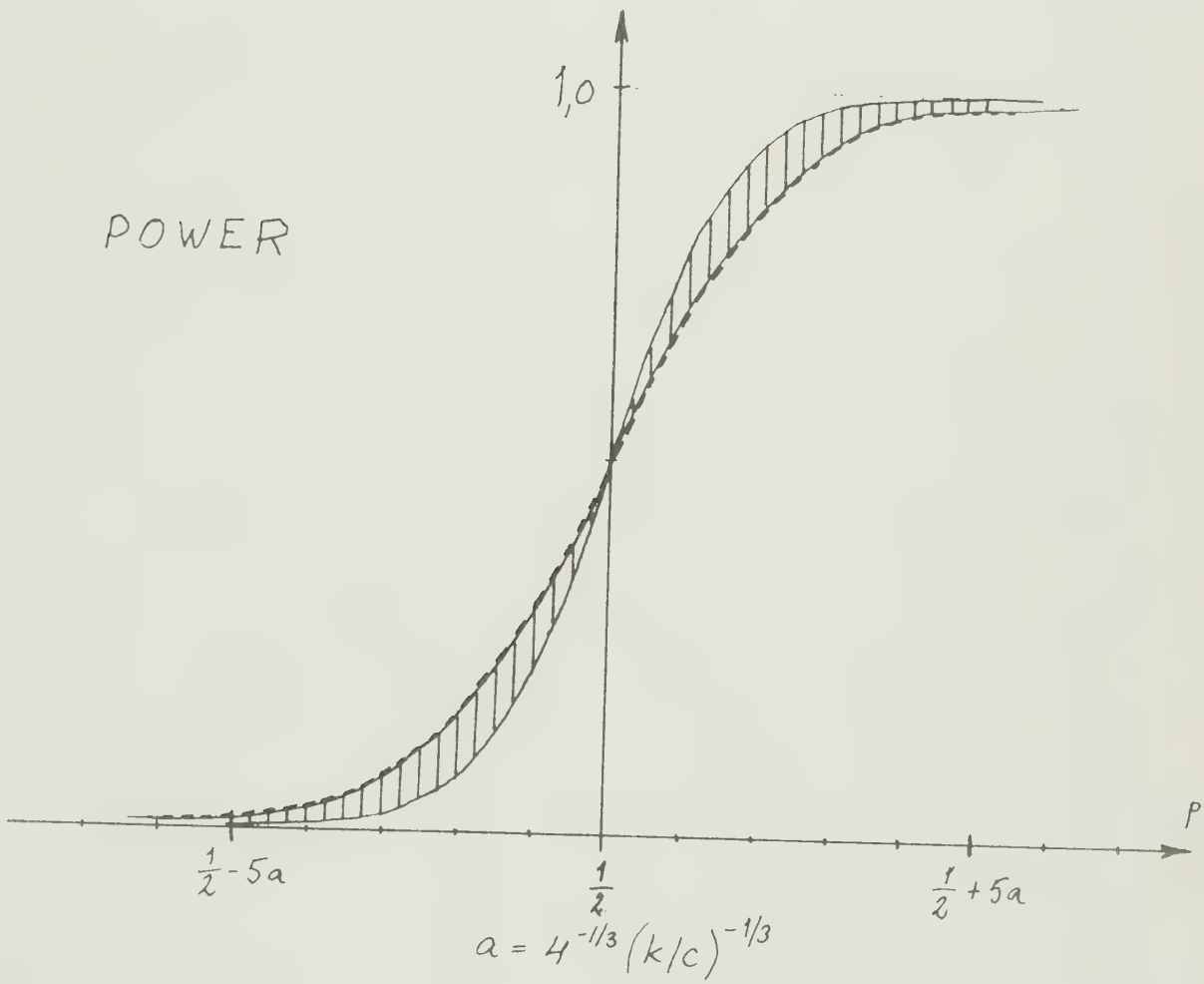
$$4^{-1/3} k^{2/3} c^{1/3} \left\{ \frac{|m|}{1 + e^{|m| \eta \sigma^2_{max}/\sigma^2}} + \right.$$

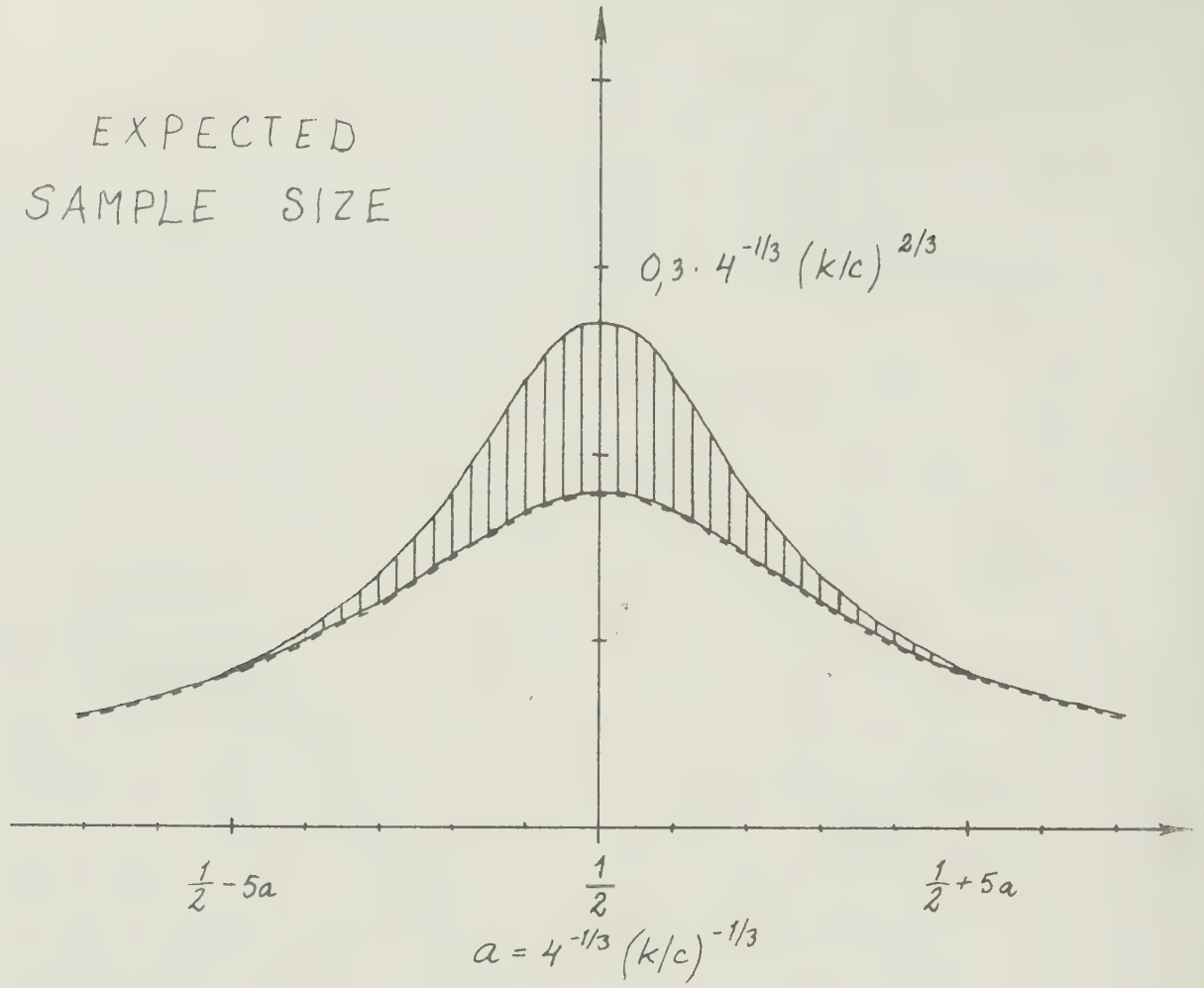
$$\left. + \frac{\eta}{2m} \frac{e^{m \eta \sigma^2_{max}/\sigma^2} - 1}{e^{m \eta \sigma^2_{max}/\sigma^2} + 1} \right\}$$

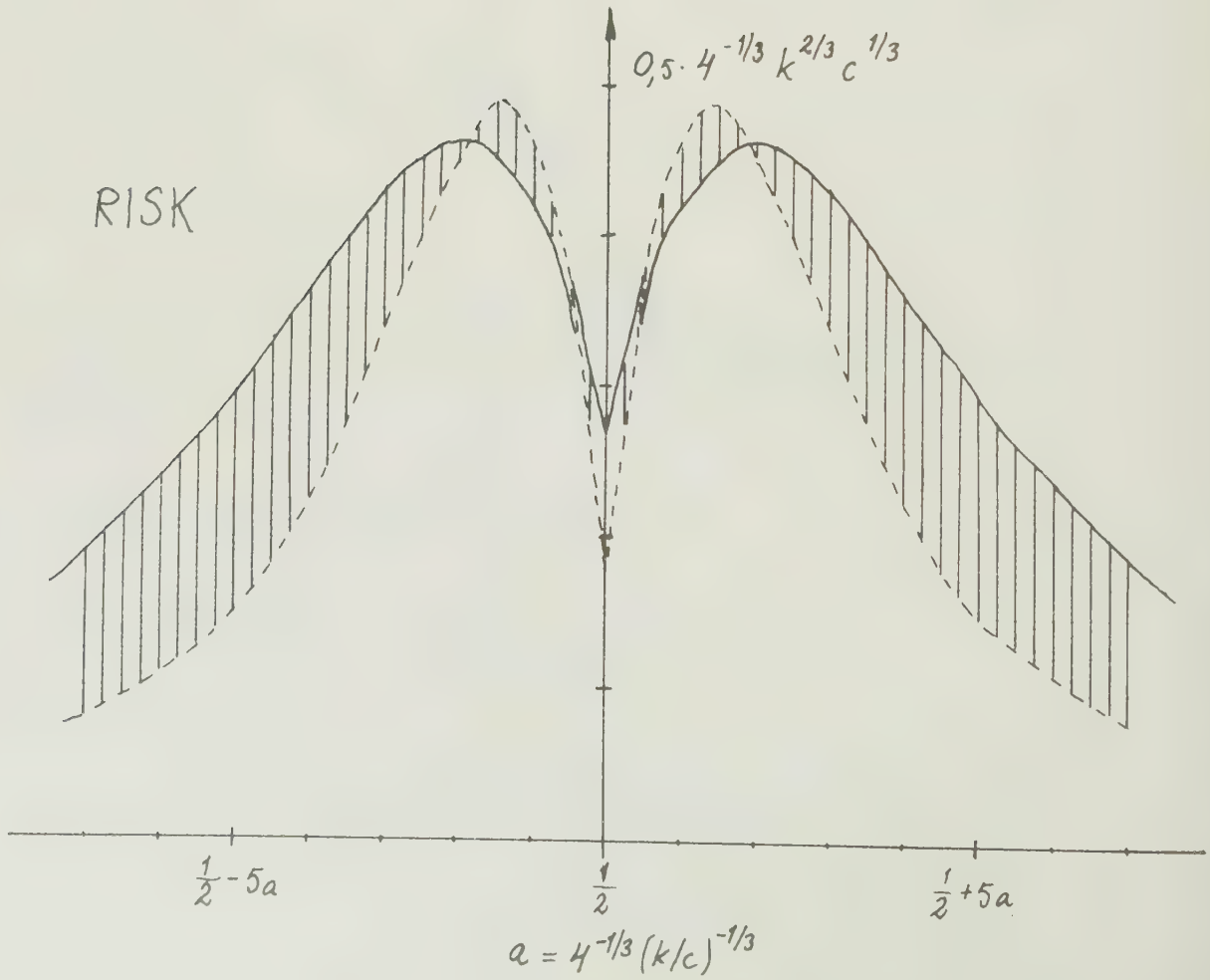
where

$$m = 4^{1/3} (k/c)^{1/3} (p-p_0).$$

In the figures are also given the asymptotically approximate power, sample size and risk of the sequential sign test in form of dotted lines.




EXPECTED
SAMPLE SIZE
$0{,}3 \cdot 4^{-1/3} (k/c)^{2/3}$
P
$\frac{1}{2} - 5a$
$\frac{1}{2}$
$\frac{1}{2} + 5a$
$a = 4^{-1/3} (k/c)^{-1/3}$


RISK
$0{,}5 \cdot 4^{-1/3} k^{2/3} c^{1/3}$
$\frac{1}{2} - 5a$
$\frac{1}{2}$
$\frac{1}{2} + 5a$
$a = 4^{-1/3} (k/c)^{-1/3}$

7. Future development

There are many statistical problems in which application of asymptotically minimax theories seem attractive and many areas of statistics in which a further develpoment of the theory could be useful. One such area is two-parameter exponential class problems, and here follows some specific problems.

In traffic statistics there is a need for good sequential methods of testing which intensity is the greatest in two Poisson processes or two more general point processes. A suitable distance parameter in this case could be the logarithm of the quotient of the intensities. Similar problems occur in tests concerning failure rates in two equipments of some kind.

In statistical quality control are often used control charts for means combined with control charts for standard deviation or range. If it is supposed that the controlled measure is normally distributed, this is a two-parameter exponential class situation, and it is reasonable in many cases to assume that the losses of making the different decisions depend only on the probability that the controlled measure will fall outside some tolerance limits. Asymptotic minimax criterias can lead to methods of statistical quality control which are completely different from those being in common use (both for the non-sequential and sequential case) and which are better when this decision-theoretic formulation is appropriate.

Another subject that could be studied further is asymptotic power and expected sample size of more general classes of sequential rank tests than those treated in the paper D. The general formulation of theorem 4.2 in Hajek [9] indicates that such results could be expected, since it enables us to generalize lemma 3 in the paper D,

although there are also other difficulties to overcome in proving weak convergence of more general rank processes.

Within the presented theory there are also things, that ought to be developed further. For practical use of the proposed tests it would be valuable to have good estimates of the maximal shortcomings of the proposed test compared to minimax tests, and this is one area that need further development. Finally of cause it would be nice to have a proof that the sequential Wilcoxon tests are asymptotically minimax, if they are.

8. Acknowledgement

I wish to thank Harald Bergström, Professor of Mathematical Statistics at Chalmers University of Technology and The University of Göteborg, for his encouragement and stimulating interest in my research work.

REFERENCES.

[1] Bell, C.B., Moser, J.M. and Thompson, R. (1966). Goodness criteria for two-sample distribution-free tests. Ann. Math. Statist.,37 133-142.

[2] Berk, R.H. and Savage I.R. (1968). The information in a rank-order and the stopping time of some associated SPRT's. Ann. Math. Statist., 39, 1661-1674.

[3] Billingsley, P. (1968). Convergence of Probability Measures, Wiley, New-York.

[4] Birnbaum, Z.W. and Klose, O. (1957). Bounds for the variance of the Mann-Whitney statistic. Ann. Math. Statist., 28, 933-945.

[5] Bradley, R.A. , Martin, D.C. and Wilcoxon, F. (1965). Sequential rank test I. Monte Carlo studies of the two-sample procedure.Technometrics, 7, 463-483.

[6] Bradley, R.A., Merchant, S.D. and Wilcoxon, F.A. (1966). Sequential rank tests II. Modified two-sample procedures. Technometrics 8, 615-623.

[7] Chapman, D.G. (1958). A comparative study of several one-sided goodness of fit tests. Ann. Math. Statist. 29, 655-674.

[8] Chernoff, H. and Savage, I.R. (1958). Asymptotic normality and efficiency of certain nonparametric tests. Ann. Math. Statist. 29, 972-994.

[9] Chernoff, H. (1959). Sequential design of experiments. Ann. Math. Statist. 30, 755-770.

[10] Chung, K.L. (1954). On a stochastic approximation mehtod. Ann. Math. Statis 25, 463-483.

[11] De Groot, M.H. (1960). Minimax sequential tests of some composite hypotheses. Ann. Math. Statist. 31, 1193-1200.

[12] Doksum, K. (1966). Asymptotically minimax distribution-free procedures. Ann. Math. Statist. 37, 619-628.

[13] Doksum, K. (1967). Asymptotically optimal statistics in some models with increasing failure rate average. Ann. Math. Statist. 38, 1731-1739.

[14] Doksum, K. (1969). Minimax results for IFRA scale alternatives. Ann. Math. Statist. 40, 1778-1783.

[15] Dvoretsky, A., Kiefer,J. and Wolfowitz, J. (1953). Sequential decision problems for problems with continuous time parameter - testing hypotheses. Ann. Math. Statist. 24, 254-264.

[16] Ghosh B.K. (1970). Sequential Tests of Statistical Hypotheses. Addison-Wesley, Reading, Mass.

[17] Gikhman, I.I. and Skorokhod, A.V. (1969). Introduction to the Theory of Random Processes. W.B. Saunders Company.

[18] Govindrajulu, Z., LeCam, L. and Raghavachari, M. (1965). Generalization of Chernoff-Savage theorems on asymptotic normality of nonparametric test statistics. Proc. Fifth Berkeley Symp. Math. Statist. Prob. 609-638.

[19] Hajek, J. (1969). Asymptotic normality of simple linear rank statistics under alternatives. Ann. Math. Statist. 40, 325-346.

[20] Hall, W.J. (1969). Embedding submartingales in Wiener processes with drift with applications to sequential analysis. J. Appl. Probability 6, 612-632.

[21] Hoeffding, W. (1951). Optimum nonparametric tests. Proc. Second Berkeley Symp. Math. Statist. Prob. 83-92.

[22] Hogg, R.V. and Craig, A.T. (1965). Introduction ot Mathematical Statistics. MacMillan, New York.

[23] Holm, S. (1971). On nonparametric asympotically minimax tests. The Annals of Math. Statistics Vol 42, No. 2.

[24] Holm, S. (1973). The asymptotic minimax character of sequential binomial and sign tests. The Annals of Statistics Vol. 1, No. 6.

[25] Holm, S. (1973). Asymptotic power and expected sample size of SPR tests in one-parameter exponential families. Mathematics and Statistics. Essays in Honour of Harald Bergström. Göteborg, Sweden.

[26] Holm, S. (1973). Sequential inversion sum tests.To appear in Skandinavian Journal of Statistics.

[27] Holm, S. (1973). Asymptotic minimax character of SPR tests in one-parameter exponential classes.To appear in Scandinavian Journal of Statistics.

[28] Kiefer, J.and Sacks, J. (1963). Asymptotically optimum sequential inference and design. Ann. Math. Statist. 34, 705-750.

[29] Lehmann, E.L. (1959). Testing Statistical Hypotheses. Wiley, New York.

[30] Raiffa, H. and Schlaifer, R. (1961). Applied Statistical Decision Theory. Harvard Univ. Press.

[31] Rao, C.R. (1965). Linear Statistical Inference and Its Applications. Wiley, New York.

[32] Renyi, A. (1970). Probability Theory, North-Holland Publ. Com., Amsterdam.

[33] Rudin, W. (1953). Principles of Mathematical Analysis. McGraw-Hill, New York.

[34] Ruist, E. (1954). Comparison of tests for nonparametric hypothesis. Ark. Mat. 3, 133-163.

[35] Savage, I.R. and Savage, L.J. (1965). Finite stopping time and finite expected stopping time. J. Roy. Statist. Soc. Ser. B. 27, 284-289.

[36] Savage, I.R. and Sathuraman, J. (1966). Stopping time of a rank-order sequential prbability test based on Lehmann alternatives. Ann. Math. Statist. 37, 1154-1160.

[37] Schwartz, G. (1962). Asymptotic shapes of Bayes sequential testing regions. Ann. Math. Statist. 33, 224-236.

[38] Whitt, W. (1970). Weak convergence of probability measures on the function space $C(0;\infty)$. Ann. Math. Statist., 41, 939-944.

[39] Wilcoxon, F., Rhodes, L.J. and Bradley, R.A. (1963). Two sequential two-sample grouped rank test with applications to screening experiments. Biomedtrics 19, 59-84.

[40] Zacks, S. (1971). The Theory of Statistical Inference. Wiley, New York.